Department of Library and Information Science

Bharathidasan University
Tiruchirappalli-620024

Name of the Programme: M.Lib.I.Sc

Course - 2.2 Information Processing and Retrieval Systems

Course Code:P21MLS7

Unit-III: Organization of digital resources – Metadata standards – Dublin core, MARC21, ISO 2709, UNIMARC, CCF and DOI (Digital Object identifier)

Course Teacher: Dr.C.RANGANATHAN
Professor
e-mail: cranganathan72@gmail.com

What is Metadata?

- Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information.

- Data about data – Digital speak for what librarians have been doing much before the Internet – surrogates, catalogs

- A metadata record consists of a set of attributes, or elements, necessary to describe the resource in question

- Structured information
  - Describes, explains, locates an information resource
  - Makes it easier to retrieve, use or manage an information resource

What Does Metadata Describe?

- Papers, articles
- Information pages
- Images
- Sound
- Collections
- User profiles
- Spatial data

Types of Metadata

- Descriptive
  - Purpose: Resource discovery and identification
  - Ex.: Title, abstract, author, URL, keywords, etc.
- Administrative & Rights Management
  - Purpose: Help manage a resource
  - Ex.: Who created and when, who can access, content format, rights information, etc.
- Structural metadata
  - Purpose: Document structure
  - Ex.: Chapter, section, paragraph

Metadata Schemes and Element Sets

- Many different metadata schemes are being developed in a variety of user environments and disciplines. Some of the most common ones are discussed in this section

- Dublin Core
- Metadata Encoding and Transmission Standard (METS)
- Metadata Object Description Schema (MODS)
- Learning Object Metadata
- The Encoded Archival Description (EAD)
- MPEG Multimedia Metadata

DUBLIN CORE

- The **Dublin Core** metadata element set is a standard for cross-domain information resource description.
- It provides a simple and standardised set of conventions for describing things online in ways that make them easier to find.
- Dublin Core is widely used to describe digital materials such as video, sound, image, text, and composite media like web pages.
- Implementations of Dublin Core typically make use of XML and are Resource Description Framework based. Dublin Core is defined by ISO in ISO Standard 15836, and NISO Standard Z39.85-2007.

Background

- The "Dublin" in the name refers to Dublin, Ohio, U.S., where the work originated from an invitational workshop (the "OCLC/NCSA Metadata Workshop") hosted in 1995 by OCLC, a library consortium that is based there. (NCSA is the National Center for Supercomputing Applications.)
- The "Core" refers to the fact that the metadata element set is a basic but expandable "core" list.

- The semantics of Dublin Core were established and are maintained by an international, cross-disciplinary group of professionals from librarianship, computer science, text encoding, museums, and other related fields of scholarship and practice.

Dublin Core Metadata Element Set (DCMES)

- consists of 15 metadata elements:
- **TITLE**: The name given to the resource by the CREATOR or PUBLISHER.
- **CREATOR**: The person(s) or organization(s) primarily responsible for the intellectual content of the resource; the author.
- **SUBJECT**: The topic of the resource; also keywords, phrases or classification descriptors that describe the subject or content of the resource.
- **DESCRIPTION**: A textual description of the content of the resource, including abstracts in the case of document-like objects; also may be a content description in the case of visual resources.
- **PUBLISHER**: The entity responsible for making the resource available in its present form, such as a publisher, university department or corporate entity.
- **CONTRIBUTORS**: Person(s) or organization(s) in addition to those specified in the CREATOR element, who have made significant intellectual contributions to the resource but on a secondary basis.

Cont..,

- **DATE**: The date the resource was made available in its present form.
- **TYPE**: The resource type, such as home page, novel, poem, working paper, technical report, essay or dictionary. It is expected that TYPE will be chosen from an enumerated list of types.
- **FORMAT**: The data representation of the resource, such as text/html, ASCII, Postscript file, executable application or JPG image. FORMAT will be assigned from enumerated lists such as registered Internet Media Types (MIME types). MIME types are defined according to the RFC2046 standard.
- **IDENTIFIER**: A string or number used to uniquely identify the resource. Examples from networked resources include URLs and URNs (when implemented).
- **SOURCE**: The work, either print or electronic, from which the resource is delivered (if applicable).
- **LANGUAGE**: The language(s) of the intellectual content of the resource.

Cont..,

- **RELATION**: The relationship to other resources. Formal specification of RELATION is currently under development.
- **COVERAGE**: The spatial locations and temporal duration characteristics of the resource. Formal specification of COVERAGE is also now being developed.
- **RIGHTS MANAGEMENT**: A link (URL or other suitable URI as appropriate) to a copyright notice, a rights-management statement or perhaps a server that would provide such information in a dynamic way.

Digital Object Identifier(DOI)

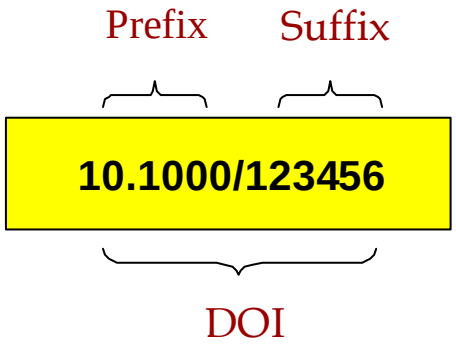
- Created in 1998;
- supports the needs of the intellectual property community in the digital environment, by the development and promotion of the Digital Object Identifier (DOI) system as a common infrastructure for content management;
- controlled by a Board, elected by the members of the Foundation;
- activities of the Foundation are controlled by its members, operating under a legal Charter and formal By-laws;
- membership is open to all organizations with an interest in electronic publishing, content distribution, rights management, and related enabling technologies;
- membership is international.

Today the foundation has over 200 companies using several million DOIs.

Digital Object Identifier (DOI)

- DOI is a system for persistent and actionable identification and interoperable exchange of intellectual property on digital networks;

- DOI is made up of two components, the prefix and the suffix;



- Value of the DOI system lies in its combination of Resolution, Metadata and Policy.

Features of DOI

Resolution -

- ensures persistence by resolving the DOI to a current associated value such as a URL;
- resolution may be to multiple pieces of data (multiple resolution);
- the Handle System is the resolution system used.

Metadata -

- based on <indecs> activity;
- consistent with systems such as ONIX and MPEG-21 rdd;
- enables mappings between application areas consistently.

Policy -

- provides rules and mechanisms for implementation;
- use of Registration agencies which operate under same rules as an operational federation.

Benefits of DOI

Actionable Identifier -

- a user can use a DOI to *do* something;
- the technology, which underlies the DOI, facilitates much more complex applications than simple location finding;
- DOI identifies the intellectual property entity itself rather than its location.

Persistent Identifier -

- if ownership of the entity or the rights in the entity change, the identification of that entity should not (and does not) change;
- the responsibility for managing the DOI changes, but not the DOI itself.

Interoperable Identifier -

- DOI System has been designed to be able to interoperate with past, present and future technologies.

OPAC-ONLINE PUBLIC ACCESS CATALOGUE

INTRODUCTION

- OPAC is a computerized library catalogue available to the public.
- It is such type of cataloguing arrangement that is considerably more user friendly than a card catalogue since it provides a variety of help to the users, and can guide the users in a step by step manner in looking for information that he is seeking. Thus, we can say that an OPAC is a computer held catalogue that is possible to be searched by a user sitting at a remote terminal.

Cont..,

- OPAC however provides for more accessibility, particularly by subject, than what is possible by a card catalogue.
- OPAC may be designed for an individual library and consulted within its building or it may be made available outside the library through Internet also.
- These days most OPACs are accessible over the Internet to users all over the world.

Reasons for Creation

- Books that contain information on a particular topic may be shelved in several in several deferent areas of the library, in special collections or even in different libraries together.
- Books are shelved in the library in call number order, so one has to be very familiar with the classification schemes used in the library to find the small area devoted to topic.

Methods of searching OPACs

- **Searching by subject**
- **Search by Author**
- **Search by Title**
- **Search by key words**

Merits of OPACs:

Most of the OPACs are very user friendly and are helpful to main users who have no idea about searching materials in the libraries.

The user after having access to such catalogues usually becomes self sufficient and seldom require the services or help of the reference staff while using libraries or accessing information.

Cont..,

- OPACs provide access to the holding of the library through different catalogues and indexes such as the author catalogue, the title catalogue, the subject catalogue, the classified catalogue, publisher's index, conference place index and KWIC/KWOC indexes.
- All possible are combination sources using Boolean operators that yield satisfying and precise results, however complex the query.

Cont..,

- OPACs also provide the facility to request acquisition of titles, to reserve materials and to send personalized SDI, overdue/recall/collect notice and messages by E-mail.

MARC

- The organization and arrangement of data elements in a particular record is called **Record format** for entering the information and to display output in a particular database.
- In this field MARC format also came into existence, which was developed by Library of Congress in 1966 as a pilot project. It was the designing of a format capable of incorporating bibliographic description for all forms of materials i.e. books, periodicals, articles etc.

MARC

- The format development was started by Library of Congress in 1965-1966 as a pilot project, known as MARC-I.
- Similar work started in the UK by the Council of the BNB as BNB MARC Project for converting BNB in machine readable format.
- These parallel developments led to Anglo-American cooperation on the MARC-II project, which was instated in 1968.

Cont..,

- MARC-II was to prove instrumental in defining the concept of MARC as a communication format for the representation and communication of bibliographic and related information in machine readable form. MARC format includes up to 61 data elements, of which 25 are directly searchable.
- Now the MARC format is well established standard for the exchange of cataloguing data.

AIM

- To communicate bibliographic and related information in machine readable form in such a way that records could be reformatted under any conceivable purpose.

Structure of MARC-II

- The format of MARC-II structure includes.
- Leader
- Record Directory
- Variable Fields

Leader

- Each MARC record contains a leader as the 24 characters and the leader provides information about the ensuing record such as the total length of the record, the type of the record code or bibliographic level.
- The type of record code specifies the form of material described in the record. The bibliographic level code denotes the relationship of the work to another bibliographic entity

Record Directory

- Record Directory shows what variable fields are in the record and their location in the record. There is a 12 characters record directory for each Variable Field.
- The existence of a record directory allows one to have variable fields in any order. The record directory also helps in the retrieval of select fields from the record.

Features of MARC-II Format

- In addition of books, the films, maps, atlases, periodicals and serials are included in it for cataloguing.
- It facilitates in magnetic tapes the documents published not only after 1969 but also the retrospective documents published before 1969, the year in which it was started.
- Databases in magnetic tapes are recorded in it and print catalogues are prepared from these magnetic tapes.

Cont..,

- It provides copies of magnetic tapes.
- The bibliographies can also be compiled with the help of the computer in this format.
- BNB also used it with the assistance of LC.
- It is also used to prepare cumulative volumes and printed cards.
- Now it is available also on on-line

COMMON COMMUNICATION FORMAT (CCF)

By

C.RANGANATHAN
Assistant Professor
Department of LIS
Bharathidasan University

Contents

- **Introduction**
- Objectives
- Principles
- Structure
  - **- Record Label**
  - **- Directory**
  - **- Data field**
  - **-Record Separator**

Introduction

The importance of information and the need for its exchange activated the innovation of a number of bibliographic record formats. One of them is Common Communication Format (CCF) which is a structure format for crating bibliographical record and for exchange of records between group of information agencies and libraries.

UNESCO had developed the Reference Manual with the help of International Council of Scientific Union/ Abstracting Board (ICSU/AB) but it had been not accepted by many organizations. These organizations continued to approach UNESCO for assistance in developing bibliographic information system.

In 1972, the UNESCO/PGI sponsored an International symposium on Bibligraphic Exchange Formats. As a result of this symposium Unesco setup the Adhoc Group for the establishment of the **Common Communication Format** (CCF).

CCF was first published in 1984. Its second edition was published in 1988 in two volumes called CCF/B and CCF/F. Bibliographic agencies around the world developed national and local formats based on the CCF.

Objectives

To provide a detailed and structured method for recording a number of mandatory and optional data elements is a computer readable bibliographic record for exchange purpose between two and more computerized systems.

It is also useful to a single bibliographic agency engaged is structuring its own format simultaneously to keeping compatibility with CCF. Non-computerized systems also can use CCF data elements.

Principles

The structure of the CCF is based on the International standard ISO 2709.

The core record consists of a small number of mandatory data elements essential to bibliographic description, identified in a standard manner.

The mandatory elements are augmented by additional optional data elements, identified in a standard manner.

Cont..,

A standard technique is used for accommodating levels relationships, and links between bibliographic entities.

It was also affirmed that CCF should be more than merely a new format. It should be based on and provide a bridge between the existing major international exchange formats, while taking into account the ISBD.

STRUCTURE

1. Record Label: Each CCF record begins with a fixed length label of 24 characters.

2. Directory: The directory is a table containing a variable number of 14 character entries. Each directory entry corresponds to an occurrence of a data field in the record. It comprises of five parts

Cont..,

TAG
3 Characters

LENTH OF DATA FIELD
4 Characters

STARTING CHARACTER POSITION
5 Characters

SEGMENT IDENTIFIER
1Character

OCCURRENCE IDENTIFIER
1Character

Directory

An example of a directory entry: 300 00333 00028910

The tag encoded in the first 3 digits is 300, signifying that the data field identifier here is a name of Person.

The next four digits are 0033 showing that the data field is of 33 characters in length.

The digit 000289 shows that the data field begins 289 characters after the base address of the data.

The digit 1 shows that this data field belongs to the second segment of the record.

The final digit 0 indicates that this is first occurrence of a data field tagged 300 in segment 1.

Data field

There are four basic components to a data field; they are indicators, subfield identifiers, subfield, field separator.

A data field, which has a single subfield, is organized as follows:

INDICATORS
2 Characters

SUBFIELD IDENTIFIER
2 Characters

SUBFIELD
Variable

FIELD SEPARATOR

1 Character

Record Separator

The last data field separator is followed by the record separator which is the final character of the record.

Utility

By using exchange format CCF, an information system converts the data in its processing format into the common format for export and each system is designed to enable the conversion of data from the common format for import into its processing format.

Thus, each system has to design only one conversion between its format and the CCF and back again

THANK YOU

Converting Bibliographic records into ISO-2709 format

By

C.RANGANATHAN
Assistant Professor
Department of LIS
Bharathidasan University

Contents

- Introduction
- File formats for bibliographic data
- ISO2709

Converting the Record into ISO2709 Format

- Objective
- Need of the project
- Limitation

- Implementation and Methodology
- Observation
- Conclusion

Introduction

Different file formats for exchange bibliographic data came into existence when peoples and organizations were investing the feasibility of producing catalogue data in machine-readable form.

Hence Exchange formats were developed in parallel with development of computers and other electronic storage devices to facilitate the transfer of bibliographic data between computer systems. Their use affects economies by reducing the duplication of effort.

Depending on the aim and objective of bibliographic data exchange different country and organization developed their own format.But there was not co-ordination between those organizations.

With the help of different organization such as Unesco, IFLA, ICSU-AB, UNISIST ISO have taken many steps towards the standardization of Bibliographic exchange format.

File formats for Exchange of bibliographic data

MARC (Machine-Readable Cataloguing)

UKMARC

MARC 21 (Formerly USMARC & CANMARC)

UNIMARC

CCF

What are standards?

Standards are documented agreements containing technical specifications or other precise criteria to be used consistently as rules, guidelines, or definitions of characteristics, to ensure that materials, products, processes and services are fit for their purpose.

Why is international standardization ?

The existence of non-harmonized standards for similar technologies in different countries or regions can contribute to so-called "technical barriers to trade". Export-minded industries have long sensed the need to agree on world standards to help rationalize the international trading process. This was the origin of the establishment of ISO.

- *Worldwide progress in trade liberalization*
- *Interpenetration of sectors*
- *Worldwide communications systems*
- *Developing countries*

ISO2709

(Format for Bibliographic Information Interchange on Magnetic Tape)

Principle and codings

Record Structure

Advantages of ISO2709

Principle and codings

A recording comprises in ISO the 2709 following parts:

- **The guide**, continuation of 24 numbered natures from 0 to 20
- **The repertory**, which comprises a variable succession of numerical natures
- **Bibliographical data** themselves

Record Structure

- Record label
- Directory
- Data fields
- Record separator

Contd-

Record Label

Each bibliographic record begins with a fixed-length label of 24 characters, the contents of which are as follows:

Character Position	Contents
0-4	Record Length (La, Di, DF, RS)
5	Record Status
6	Blank
7	Bibliographic level
8	Blank
9	Blank

Character Position	Contents
10	‘2’ Indicator length
11	‘2’Subfield identifier length
12-16	Base address of data
17-19	Blank
20	‘4’Length of datafield
21	‘5’ Starting character position
22	‘2’ Segment identi & Occur identi
23	Blank

Directory

- **Tag**
  - **Length**
  - **Starting character position**
  - **Segment identifier**
  - **Occurrence identifier**

Tag

A three character code identifying the data field which corresponds to directory entry

Length of datafield

A four-digit number showing how many characters are occupied the datafield, including indicators and datafield separator but excluding the record separator code if the data field is the last field in the record.

Starting Character Position

A five-digit number giving the position of the first character of the datafield relative to the base address of data, i.e. the first character of the first of the datafield

Segment Identifier

A single character (chosen from 0-9 and/or A-Z) which designates the datafield as being a member of particular segment.

Occurrence Identifier

A single character (chosen from 0-9 and A-Z) which differentiates multiple occurrences of the datafields that carry the same tag within the same record segment

Datafields

*** Indicators:** Further information of datafield

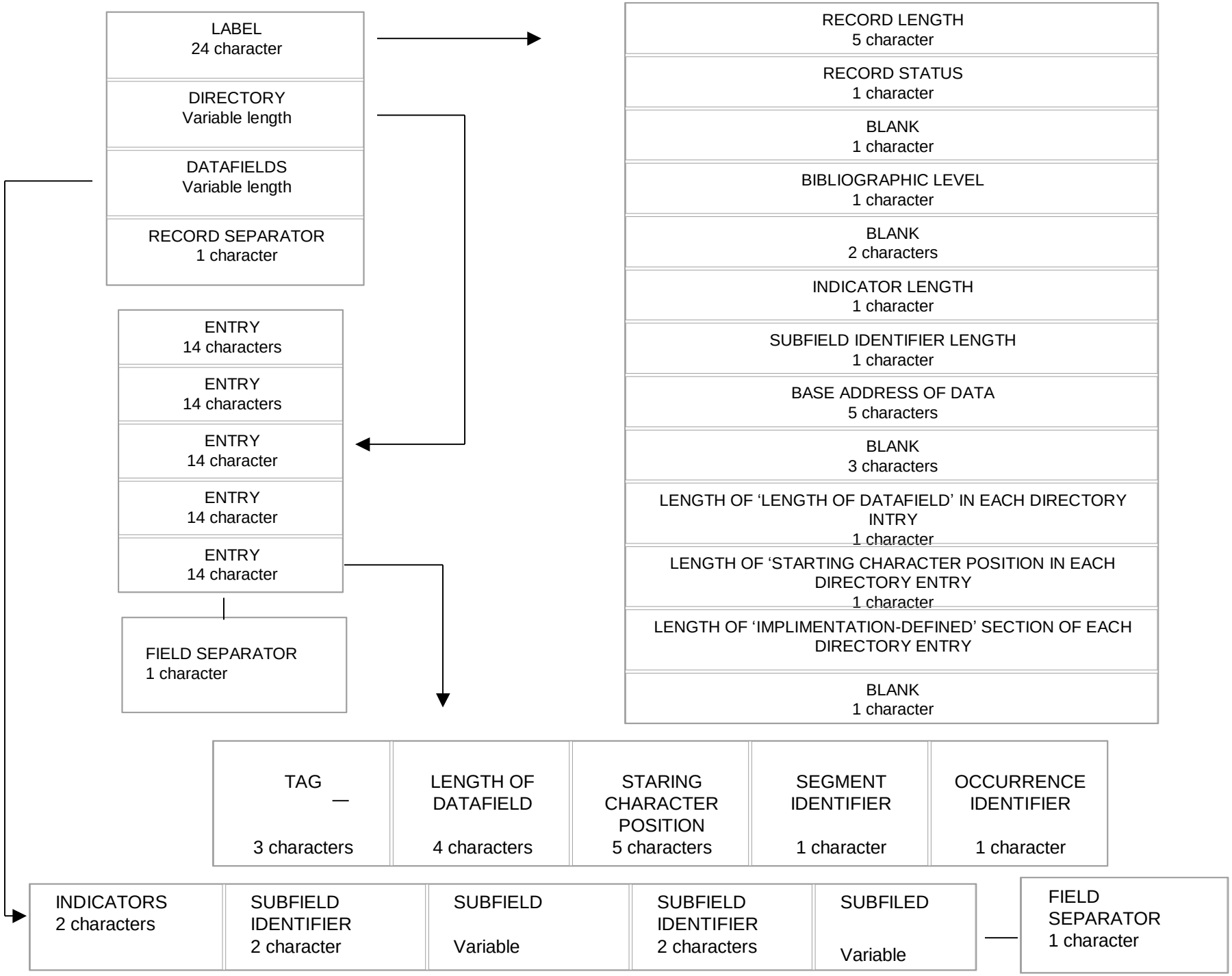
*** One or more subfields each of which is preceded by a subfield identifier**

*** A datafield separator**

Record Separator

The record separator is the final character of the record. It follows the field separator of the final datafield of the record.

DIGRAMMATIC REPRESENTATION OF RECORD STRUCTURE


LABEL
24 character
DIRECTORY
Variable length
DATAFIELDS
Variable length
RECORD SEPARATOR
1 character
ENTRY
14 characters
ENTRY
14 characters
ENTRY
14 character
ENTRY
14 character
ENTRY
14 character
FIELD SEPARATOR
1 character
RECORD LENGTH
5 character
RECORD STATUS
1 character
BLANK
1 character
BIBLIOGRAPHIC LEVEL
1 character
BLANK
2 characters
INDICATOR LENGTH
1 character
SUBFIELD IDENTIFIER LENGTH
1 character
BASE ADDRESS OF DATA
5 characters
BLANK
3 characters
LENGTH OF 'LENGTH OF DATAFIELD' IN EACH DIRECTORY INTRY
1 character
LENGTH OF 'STARTING CHARACTER POSITION IN EACH DIRECTORY ENTRY
1 character
LENGTH OF 'IMPLIMENTATION-DEFINED' SECTION OF EACH DIRECTORY ENTRY
BLANK
1 character
TAG
3 characters
LENGTH OF DATAFIELD
4 characters
STARING CHARACTER POSITION
5 characters
SEGMENT IDENTIFIER
1 character
OCCURRENCE IDENTIFIER
1 character
INDICATORS
2 characters
SUBFIELD IDENTIFIER
2 character
SUBFIELD
Variable
SUBFIELD IDENTIFIER
2 characters
SUBFILED
Variable
FIELD SEPARATOR
1 character

Advantages of ISO2709

* It provides a small number of mandatory data elements, which are recognized by all sectors of the information community as essential in order to identify an item.

* It gives mandatory data elements that are sufficiently flexible to accommodate varying descriptive practices.

CONT…

CONT…

* **It also provides a number of optional elements, which may be useful to describe an item according to practices of the agency, which creates the record.**

* **It provides a mechanism for linking records and segments of records without imposing on the originating agency any uniform practice regarding the treatment of related groups of records or data elements**

Conversion of Record into ISO2709

- Develop our own S/W
- Use existing S/W

  e.g. Fangorn, TXT2ISO

Fangorn:

Hugo Besemer,

Bredeweg 17,6668 AR Randwijk,

the Netherlands, e-mail BESEMER@JKA.WAU.NL .

TXT2ISO:

L.J Haravu (Manager),

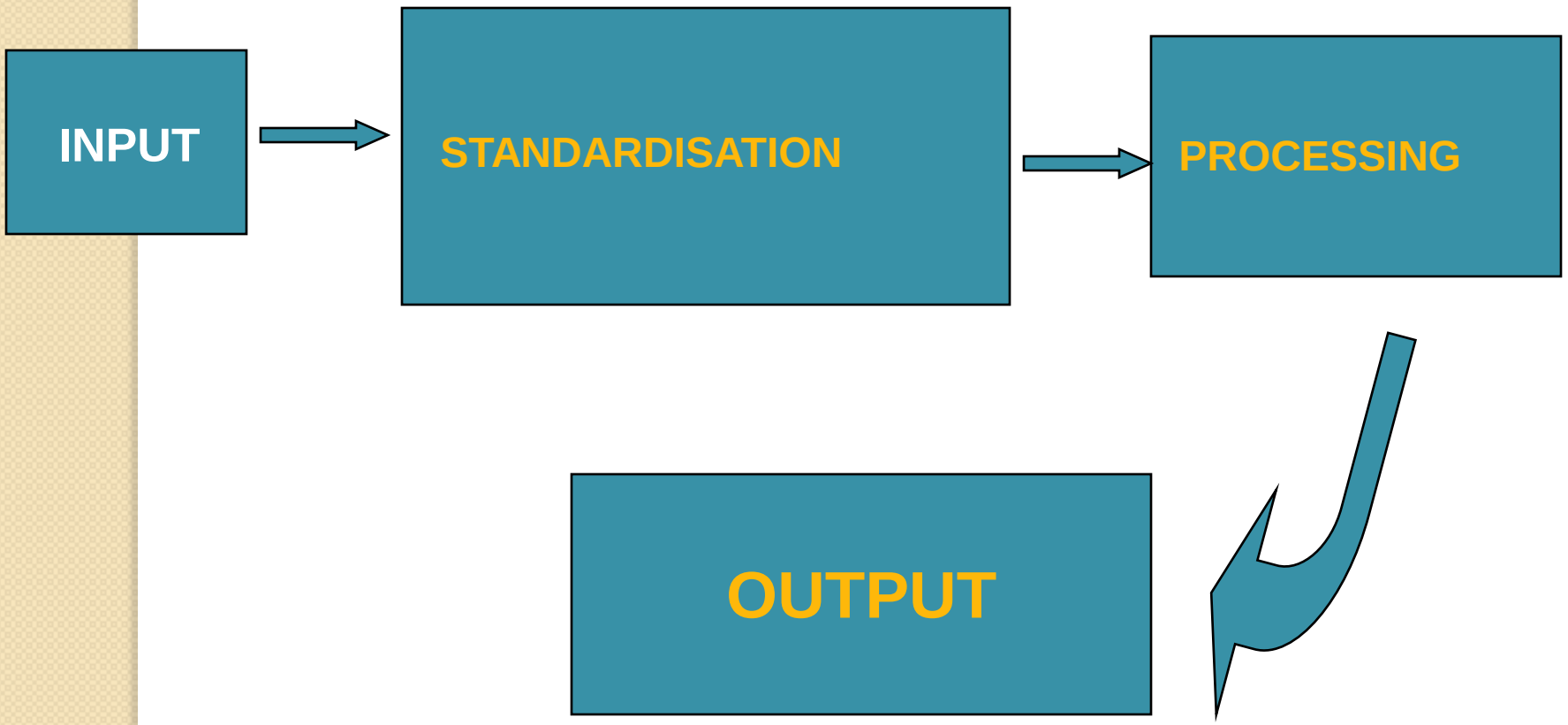
Library & Documentation Services

ICRISAT, Patancheru P.O.

Andhra Pradesh 502 324, India.

Demo

- Input file
- Standardization of text file using "C"
- Running the Utility
- Output


INPUT
STANDARDISATION
PROCESSING
OUTPUT

Conversion into Bibliographic data

File import into WINISIS

OBSERVATION AND CONCLUSION

The program works only with current content data. The program can be generalized so that it is able to convert records from different databases into the file format as required by the "TXT2ISO.EXE" program. Numbers of researchers come to NCSI to avail different information services. They are provided with bibliographic data in text format. Program can be modified so that these records given to users can be converted to ISO2709 format so that they can import these records into a database management system like CDS/ISIS.

THANK YOU

UNIMARC (UNIVERSAL MACHINE READABLE CATAALOGUE)

- International exchange of MARC data, IFLA established a Working Group on Content Designators in 1972. The format developed by the Working Group became UNIMARC.

Purposes

- The purpose of UNIMARC is to facilitate the description, retrieval and control of bibliographic items for data exchange and also for local bibliographic format. There are a number of factors that shaped UNIMARC into the flexible data package [Morataza, 1996].

STRUCTURE

- Record Structure
- ISBD
- Textual and Non-textual Material
- Multiple levels
- Parts of Items/Linking Technique

UNIMARC Functional Blocks

- 0 — Identification block, contains those numbers that identify the record of the work (e.g., ISBN, ISSN).
- 1 — Coded Information block, contains fixed length coded data elements describing various aspects of the work.
- 2 — Descriptive block, contains the areas covered by the ISBD (i.e. title, edition, imprint, collation, series) with the exception of standard number and notes.
- 3 — Notes block, contains free text statements describing various aspects of the work.
- 4 — Linking Entry block, contains standard links in numeric and textual form to other records.
- 5 — Related Title block, contains titles to be used as access points.
- 6 — Subject Analysis block, contains subject identification (e.g., UDC, Library of Congress Subject Headings, etc.).Personal and corporate names used as subjects will appear in this block.
- 7 — Intellectual Responsibility block, contains names of persons and corporate bodies responsible for the creation of the work described in access point form.
- 8 — International Use block, contains internationally agreed field that do not fit in the preceding blocks, 0– to 7–. This block includes fields on originating agency, ISSN Center, general cataloguer's note and electronic location and access.
- 9 — National Use block, reserved for national use by agencies where UNIMARC is the basis of the domestic format.